Pollution and city size: can cities be too small?*

Rainald Borck†
University of Potsdam, CESifo and DIW Berlin

Takatoshi Tabuchi‡
University of Tokyo and RIETI

July 2017, revised March 13, 2018

Abstract

We study optimal and equilibrium size of cities in a city system model with pollution. Pollution is a function of population size. If pollution is local or per capita pollution increases with population, equilibrium cities are too large under symmetry; with asymmetric cities, the largest cities are too large and the smallest too small. When pollution is global and per capita pollution declines with city size, cities may be too small under symmetry; with asymmetric cities, the largest cities are too small and the smallest too large if the marginal damage of pollution is large enough. We calibrate the model to US cities and find that the largest cities may be undersized by 3-4%.

JEL classification: R12, Q54

Keywords: optimal city size distribution, agglomeration, pollution

*We thank two referees, Stefan Bauernschuster, Jan Brueckner, M. Morikawa, T. Morita, Michael Pflüger, and M. Yano as well as participants at MCC Berlin, Tinbergen Institute, University Duisburg-Essen, Free University Berlin, in Minneapolis (UEA), Lisbon (EMUEA), Münster and Dresden (VFS), Tokyo (RIETI), Osaka (spatial economics conference) and Pau (Location choices and environmental economics workshop) for comments and suggestions. The first author thanks the German Science Foundation (DFG) and the second author thanks RIETI for financial support.

†University of Potsdam, Faculty of Economic and Social Sciences, August-Bebel-Str. 89, 14482 Potsdam, Germany, e-mail: `rainald.borck@uni-potsdam.de`.

‡University of Tokyo, Faculty of Economics, Hongo 7-3-1, Bunkyo-ku, Tokyo, Japan, e-mail: `ttabuchi@e.u-tokyo.ac.jp`

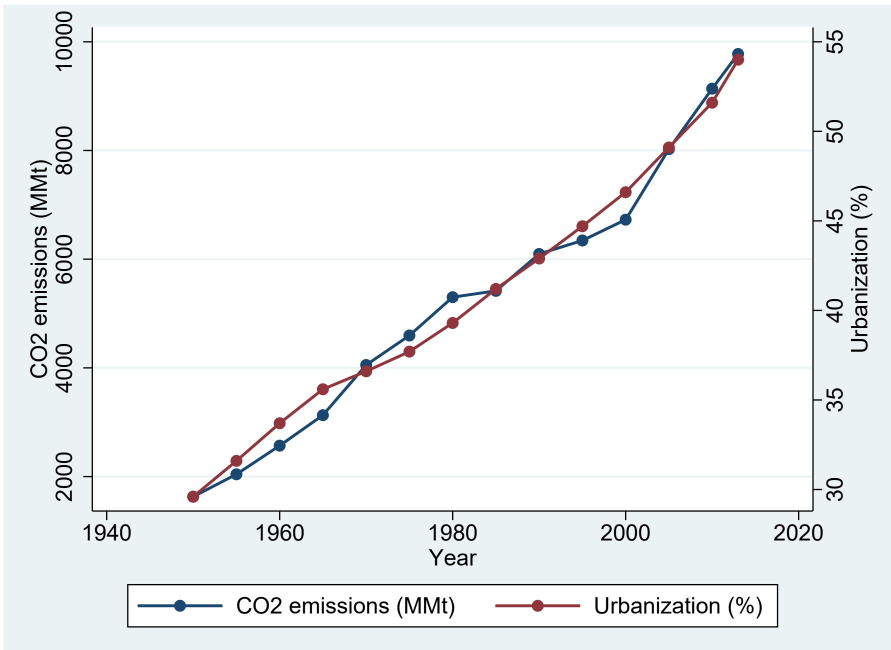

Figure 1: World urbanization and CO_2 emissions 1950-2013
Source: UN Population Division, CDIAC

1 Introduction

Urbanization is rapidly increasing, especially in developing countries. According to the UN Population Division, urbanization worldwide will increase from 51.6% in 2010 to 66.4% in 2050, and from 46.1% to 63.4% in the developing world. Some commentators are afraid that this urbanization may have adverse environmental consequences. For instance, Seto *et al.* (2012) argue that the projected urbanization until 2030 leads to significant loss of biodiversity and increased CO_2 emissions due to deforestation and land use changes. Urban economic activities such as manufacturing production, commuting, and residential energy use also contribute to pollution. Fig. 1 shows that over the last half century, urbanisation and CO_2 emissions have moved together. Of course, this may not be a causal relation.

In fact, some writers who claim that large, densely populated cities produce lower per capita emissions. Glaeser and Kahn (2010) show that in the US, inhabitants of large, densely populated cities such as New York City and San Francisco tend to produce lower CO_2 emissions from transport and residential energy use than those living in smaller and less densely populated cities, controlling for factors such as local weather. Glaeser (2011) writes about this *Triumph of the City* and in the subtitle succinctly states: "How our greatest invention makes us richer, smarter, *greener*, healthier, and happier" (our emphasis). This line of reasoning has prompted organizations such as the OECD and the World Bank to advocate high density urban development to mitigate environmental pollution.

Therefore, an important policy question is whether big cities are good or bad for the environment, especially in developing countries such as China, where new cities are springing up by the minute. While on the one hand, migrants flock to cities to take advantage of their economic opportunities, on the other hand, concern about congestion, environmental pollution and other side effects is mounting. So what is the optimal size of cities that are affected by environmental pollution? And what would be the unregulated equilibrium city size?

In this paper, we build a simple model of a city system to study how the equilibrium and optimal city size distributions are affected by environmental pollution. We use a standard monocentric city model, where people work, consume goods and housing in cities. Agglomeration externalities make workers more productive in big cities. Pollution is related to city size since it is a by-product of urban production, commuting and housing. In line with reality, we assume that externalities arising from pollution are not internalized. We distinguish between pollution which is purely local, such as certain kinds of emissions from traffic, and pollution which spills over between cities, such as greenhouse gas (GHG) emissions. When cities are symmetric, we find that with local pollution, equilibrium cities are too large and there are too few of them, mirroring the classic result of Henderson (1974). By contrast, when pollution is global, we find that equilibrium cities may be either too small or too big. The former case can occur when per capita pollution falls with city size. We also study the model with a given number of asymmetric cities. With local pollution, we find that the largest cities are too large and the smallest cities too small. With global pollution, if per capita pollution decreases with city size and the marginal damage of pollution is large enough, the largest cities are too small and the smallest too large.

We also quantify the extent to which cities may be undersized, using a calibrated version of the model. We use some standard parameter values from the literature, and, using data from Fragkias et al. (2013), we estimate the effect of the size of US metropolitan areas on CO_2 emissions. We find that doubling city size reduces per capita CO_2 emissions by 13 percent. In the symmetric city case, we find that cities might be undersized by up to 9 percent if pollution is global. With asymmetric cities, in the case of global pollution, the largest cities may be undersized by 3-4 percent while the smallest cities are oversized by 5-10 percent. If pollution is local, the largest cities are oversized, but by only about 0.3%.

Finally, we use an estimate of the degree of pollution spillovers (so pollution is neither completely global nor completely local) and find that the largest city is undersized by 2.4% and the smallest is oversized by 6.6%.

Our paper is related to several strands of literature. First, the literature on city systems has studied equilibrium and optimal city sizes. Henderson (1974) first showed that in equilibrium, cities are too big. This finding also comes out of the models by Tolley (1974), Arnott (1979), and Abdel-Rahman (1988). Tolley (1974) considers local pollution and actually argues that it leads to cities being too big. We show that this argument depends on whether pollution spills over to other cities and how per capita emissions change with city size. Abdel-Rahman and Anas (2004) review this literature and also discuss the role of externalities in city system models (though not externalities arising from pollution).

On the other hand, some recent papers show that cities may be too small in equilibrium. Albouy et al. (2016) show that large cities may be too small due to federal taxation, local politics, wedges due to land ownership, and the interaction between these. Eeckhout and Guner (2017) also show that spatially uniform taxation may lead to large cities being undersized, and that the optimum spatial tax system taxes individuals in large cities less than the current US tax system.[1] Like Albouy et al. (2016) and Eeckhout and Guner (2017), we show that cities may be too small.[2] While these papers have found that cities may be too small, we differ from them by analyzing equilibrium and optimal city size in a city system model with pollution. Hence, the mechanism generating the divergence of optimum and equilibrium city sizes is different. Since pollution is high on the agenda of policy makers worldwide, we think this is an important topic.

Second, there is a small but growing literature on cities and the environment more general. Related to this paper, Gaigné *et al.* (2012) and Borck and Pflüger (2015) study the interaction of agglomeration, pollution and welfare in models with a given number (two) of cities. The upshot from these papers is that pollution may rise or fall when population density increases or the city system becomes more agglomerated. They differ from our paper in that they theoretically show how pollution is affected by the location of mobile factors between two cities. They focus on the links between city size and pollution from commuting, goods transport, production and residential energy use. In a sense, they study in detail the link between city size and pollution, using microfounded models of

[1] Au and Henderson (2006) show that many Chinese are too small due to the migration restrictions of the *hukou* system.

[2] See also Hsieh and Moretti (2017), who show that local housing supply inefficiently restricts migration of workers to high productivity cities such as San Francisco and San Jose.

urban structure and economic geography, whereas we, instead, model this link as a reduced form relation in our theoretical model, and later estimate it as a basis for our numerical simulation.

There are also some theoretical papers on urban structure and pollution, see Borck (2016), Borck and Brueckner (2016), Dascher (2014), Larson *et al.* (2012) and Tscharaktschiew and Hirte (2010). These papers, different from ours, study the interaction of urban structure and pollution within cities. Finally, Glaeser and Kahn (2010) and Larson and Yezer (2015) study empirically the relation between GHG emissions or energy use and city structure. Glaeser and Kahn (2010) find that large, dense cities in the US produce fewer GHG emissions. Morikawa (2013) finds that dense cities in Japan produce lower per capita energy consumption in the service sector, and Blaudin de Thé and Lafourcade (2016) show that residents of low density suburban households in France use more gasoline for driving. Larson and Yezer (2015) study the effect of city size on energy use in a simulation model, finding that per capita energy use does not change with city size. A number of papers from other disciplines than economics also study the relation between city size and pollution empirically, with different results, see e.g. the study of Fragkias et al. (2013) using panel data from US cities and Sarzynski (2012) who use a sample of 8038 cities world-wide in 2005.[3] Our paper is also concerned with the relation between pollution and city size, which is essential for the comparison of equilibrium and optimal city systems.

Finally, the paper is related to a growing literature on equilibrium models with either exogenous or endogenous amenities, see, e.g., Diamond (2016).[4]

We proceed as follows. The next section introduces the model of a symmetric city system. Section 3 presents the modeling of pollution. In section 4, we study the equilibrium and optimum size of cities with local and global pollution. Section 6 contains a numerical simulation, to get a sense of the possible divergence of optimum and equilibrium city size. In Section 5, we extend both the analytical and simulation results to the realistic case of asymmetric cities. The last section concludes.

2 The model with symmetric cities

There are m cities in the economy, whose total population is exogenous and denoted by N.[5] For now, we assume cities to be identical. The population size in each city is endogenous

[3] See Section 3 for more on this literature.

[4] Redding and Rossi-Hansberg (2017) survey related quantitative spatial economics models.

[5] In contrast to Albouy et al. (2016), we don't consider a rural sector in the economy.

and given by $n = N/m$. For simplicity, the city space is linear with unit width and the central business district (CBD) is a spaceless point located at $x = 0$, while the endogenous city border is denoted $\bar{x}$ (we focus on the right side of the city for simplicity). All individuals commute to the CBD and have identical preference given by

$$u(s, z, E) = s^{\alpha} z^{1-\alpha} E^{-\beta}, \tag{1}$$

and the budget constraint is

$$w = z + rs + tx, \tag{2}$$

where s is housing floor space (equivalently land consumption), z is consumption of a composite non-housing good, E is pollution, w is wage income, r is the housing rent per square meter, t is the commuting cost per mile, x is distance from the CBD, and $0 < \alpha < 1$, and $\beta > 0$.

Consumers choose s and z to maximize (1) subject to (2). From this we get optimal housing consumption

$$s(w - tx, r) = \frac{\alpha(w - tx)}{r}. \tag{3}$$

Consumers are mobile within and between cities, and land is rented to the highest bidder. We can now solve for households' bid rent, i.e., the maximum amount the household would be willing to pay per unit of land. Using (3) and (2) in (1) and solving $u(z, s, E) = \overline{u}$ gives

$$r(w - tx, E, v) = (w - tx)^{1/\alpha} E^{-\beta/\alpha} v^{-1/\alpha}, \tag{4}$$

where $v \equiv \alpha^{-\alpha} (1 - \alpha)^{-(1-\alpha)} \overline{u}$.

The two equilibrium conditions in the representative city are:

$$r(w - t\bar{x}, E, v) = r_A \tag{5}$$

$$\int_0^{\bar{x}} \frac{1}{s(w - tx, E, v)} dx = n, \tag{6}$$

where r_A is the agricultural land rent. Eq. (5) states that at the city border, land rent just equals the agricultural land rent. Eq. (6) says that the population n fits into the city between 0 and $\bar{x}$.

Suppose that there are external economies of scale at the city level, for instance because of gains from individual specialization. Total city production is assumed to be $Y = n^{1+\gamma}$,

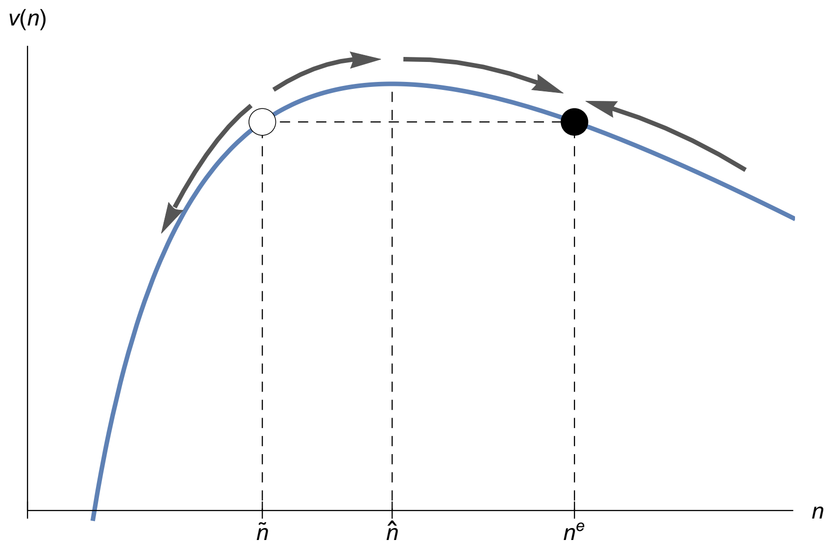

Figure 2: Equilibrium and optimal city size with local pollution

Note: The figure shows optimal ($\hat{n}$) and possible equilibrium (n^e) city size with local pollution

with $\alpha > \gamma > 0$ and the individual wage is $w = n^\gamma$.[6] The restriction $\alpha > \gamma$ is necessary to ensure that in the absence of pollution, utility is an inversely U-shaped function of city size, as shown in Fig. 2, so that the optimal city size without pollution is finite (see eq. (11)).

Substituting (3) and (4) into (5) and (6) and solving gives the city border and indirect utility

$$\bar{x} = \frac{n^\gamma \left[1 - r_A^\alpha (r_A + tn)^{-\alpha}\right]}{t} \tag{7}$$

$$v = n^\gamma (r_A + tn)^{-\alpha} E^{-\beta}. \tag{8}$$

Eq. (7) shows that the city expands as population grows. It shows that $\bar{x}$ is not directly affected by pollution although it is indirectly affected by population change through intercity migration. Eq. (8) shows the standard tradeoff induced by an increasing city population: on the one hand, utility increases with n due to agglomeration forces, on the other hand, it decreases because of longer commutes and competition for land, which results in higher land rents. In the next section, we model pollution in order to study how it affects this fundamental tradeoff. In particular, the pertinent question is how reallocating population among cities affects the disutility from pollution.

[6] Duranton and Puga (2004) show that several different mechanisms lead to the same functional form, such as gains from specialization, matching, sharing intermediate inputs, or learning.

Note that we have assumed that land is owned by absentee landowners. As is well known, efficiency analysis requires returning differential land rents to city residents. We show in Appendix B, however, that our results hold qualitatively if land is owned by city residents.[7]

3 Pollution

Pollution in city i is given by

$$E_i(\mathbf{n}) = e(n_i) + \delta \sum_{j=1,\, j\neq i}^{m} e(n_j), \quad i = 1, ..., m,$$

where $e(n_i)$ are local emissions and $0 \leq \delta \leq 1$ measures the degree of pollution spillovers. When $\delta = 0$, pollution is purely local (for instance, some forms of particulate pollution which do not diffuse over long distances). Conversely, when $\delta = 1$, pollution is purely global from the view of our city system, as is the case, for instance, for GHG emissions. Importantly, in the latter case, the environmental externality is independent of the individual's location.

An important issue in the coming analysis will be the relationship between emissions and city population, as captured by the function $e(n_i)$. We assume that city population affects local emissions through residents' economic activities, such as commuting, housing, and consumption of other goods whose production causes emissions. What do we know about this relation? In Section 6, we will try to estimate the population elasticity of pollution empirically, but here we briefly discuss theoretical an empirical studies that address this issue.

Borck and Pflüger (2015) present a theoretical model in which urban pollution is driven by commuting, residential energy use, industrial and agricultural production, and goods transport. They show that per capita pollution from industrial production and residential energy use decreases with city size, while pollution from commuting and goods transport increases. The total effect of city population on urban pollution is ambiguous and depends on parameters.

Some authors have estimated the relation between pollution and city population (or population density) empirically. Most of these papers estimate an equation of the form

[7] Albouy et al. (2016) study cross-city externalities arising from landownership requirements on migrants.

$e = Bn^{\theta}$, which we will also do in Section 6.[8] Lamsal *et al.* (2013) use cross-sectional cross-country data on NO_2 and NO_X pollution and find that the elasticity of pollution with respect to population density lies between 0.4 and 0.67. Gudipudi *et al.* (2016) study the effect of population density on CO_2 emissions and find an elasticity around 0.6, so doubling population density would reduce per-capita emissions by 24%. Fragkias et al. (2013) also estimate the effect of population on CO_2 emissions, using panel data from US cities. They find an elasticity of emissions with respect to population of 0.93. Rybski *et al.* (2016) conduct a meta-analysis of published articles that study CO_2 emissions and city size, and find that in developed countries per capita emissions decrease with city size while in developing countries per capita emissions increase with city population. However, most of these estimates seem problematic. For instance, Lamsal *et al.* (2013) use cross-sectional OLS regressions to estimate the population elasticity of pollution. But this ignores potential confounders that are correlated with population density and pollution. Fragkias et al. (2013) use panel data, but they estimate the model with random effects, which assumes that any unobserved time-invariant heterogeneity between cities is not correlated with pollution. In Section 6, we present an alternative estimate of the population elasticity of CO_2 pollution, using the same dataset as Fragkias et al. (2013).

4 Equilibrium and optimum number and size of cities

The equilibrium city size in the city system is defined by the solution of $v_i = v^*$ for all i. We focus on symmetric cities. Further, we require the equilibrium to be stable, which implies $\partial v(n)/\partial n < 0$. To study optimal city size, we assume a central planner who maximizes aggregate welfare $mnv(n)$ with respect to n and m. Using $mn = N$, this is equivalent to maximizing $v(n)$ with respect to n. Note that, from (8) follows $v(0) = 0$ so no one would ever want not to live in a city.

[8]Some papers not reviewed here estimate other functional forms, where, for instance, pollution is assumed to be a quadratic function of population.

4.1 Local pollution

Suppose first that pollution is entirely local, i.e. $\delta = 0$. Then migration is governed by the following utility differential

$$v(n_i) - v(n_j) = \hat{v}(n_i)e(n_i)^{-\beta} - \hat{v}(n_j)e(n_j)^{-\beta}, \qquad (9)$$
$$\text{where } \hat{v}(n_i) \equiv n_i^{\gamma}(r_A + tn_i)^{-\alpha},$$

and optimum city size maximizes $v(n_i) = \hat{v}(n_i)e(n_i)^{-\beta}$. We will assume that both $v(n)$ and $\hat{v}(n)$ are quasi-concave, which holds (in the neighborhood of the equilibrium and social optimum) for the parameter values used in our numerical simulations. Moreover, we assume that locally produced pollution $e(n_i)$ satisfies $e(0) = 0$ and $de/dn_i > 0$.

Since $v(n_i)$ can be shown to be inverted U-shaped, we get the standard result that equilibrium cities are too large, as in Henderson (1974). This can be seen by looking at Fig. 2. The figure shows the optimal city size $\hat{n}$ and two potential equilibrium city sizes $\tilde{n}$ and n^e.[9] Any equilibrium with city size $\tilde{n} < \hat{n}$ is unstable: if the city population were to deviate slightly from $\tilde{n}$, migration in or out of the city would occur, as indicated by the arrows. Conversely, any equilibrium with $n^e > \hat{n}$ is stable: as indicated by the arrows, a deviation from n^e would induce migration flows which restore the equilibrium. Therefore, there is a continuum of equilibria with $n^e > \hat{n}$ where $\hat{n} = n^*$ maximizes $v(n_i)$. We summarize this as:

Proposition 1 *If cities are symmetric and pollution is purely local, cities are too large in equilibrium.*

The economic intuition for this result is that in a stable equilibrium, all cities are on the decreasing part of the indirect utility curve so that negative externalities dominate at the margin. Since migrants fail to internalize the consequences of their location choices, cities are too large in equilibrium.

4.2 Global pollution

Now, let $\delta = 1$ so that pollution is global from the viewpoint of the economy. Since pollution is global, we can drop the index i from pollution E_i and write the utility difference of living

[9] Note that there is a continuum of equilibria, so all that can be said in general is that $n^e > \hat{n}$, but the exact location of the equilibrium is indeterminate.

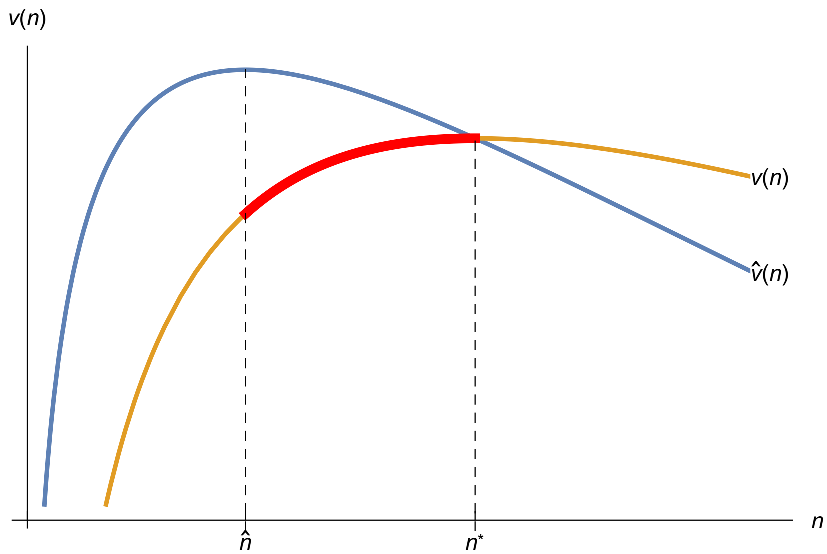

Figure 3: Equilibrium and optimum city size with global pollution
Note: The figure shows optimal (n^*) and possible equilibrium (some $n > \hat{n}$) city size with global pollution. If equilibrium city size is on the red portion of the $v(n)$ curve, it is smaller than optimal.

in city i versus j as

$$v(n_i) - v(n_j) = E^{-\beta}\left(\hat{v}(n_i) - \hat{v}(n_j)\right). \tag{10}$$

For $E > 0$, the individual migration decision is determined by the difference $\hat{v}(n_i) - \hat{v}(n_j)$, so global pollution does not affect migration decisions. Let $\hat{n}$ denote the city size which solves $\max_n \hat{v}(n)$. Setting $\hat{v}'(n) = 0$ and solving gives

$$\hat{n} = \frac{\gamma r_A}{(\alpha - \gamma)t}. \tag{11}$$

Then, by the same argument as in Henderson (1974), there is a continuum of stable equilibria with city sizes $n^e > \hat{n}$. Fig. 3 shows possible equilibrium city sizes. As before, any equilibrium with $n^e > \hat{n}$ is stable.

The optimum city size n^* is found by maximizing $v(n) = \hat{v}(n)E(n)^{-\beta}$. The first order condition can be written

$$\frac{\hat{v}'(n)n}{\hat{v}(n)} = \beta\frac{E'(n)n}{E(n)}. \tag{12}$$

At the optimum, the elasticity of (private utility) with respect to population size should equal the elasticity of total emissions, multiplied by the marginal damage of emissions. We know that $n^e \geq \hat{n}$ and that $\hat{n}$ maximizes $\hat{v}(n)$. Since $\beta > 0, E(n) > 0$ and $\hat{v}(n)$ is quasi-concave, evaluating (12) at $\hat{n}$ implies that $n^* < \hat{n}$ if $E'(\hat{n}) > 0$. Since $E(n) = m \cdot e(n) =$

$\frac{N}{n}e(n)$, we find cities are definitely too large if per capita pollution is increasing in city size. Intuitively, in this case making cities larger increases pollution, which increases the disutility from pollution. This reinforces the argument in Henderson-style models which make cities too large.

However, if per capita emissions are decreasing in city size, we find $n^* > \hat{n}$. This opens up the possibility that in equilibrium, cities may be too small. However, since there is a continuum of equilibria with $n^e > \hat{n}$, cities may also be too large. Summarizing this discussion, we have:

Proposition 2 *Suppose that pollution is global, i.e. $\delta = 1$. If per capita emissions increase with n, cities are too large in equilibrium. However, if per capita emissions decrease with n, cities may be either too small or too large in equilibrium.*

Fig. 3 illustrates the case where pollution is global and per capita emissions are decreasing with city size. The equilibrium city size is some $n^e > \hat{n}$, where $\hat{n}$ is the maximum of the function $\hat{v}(n)$. The optimum city size n^* is the maximum of the $v(n)$ curve.[10] The thick (red) part of the $v(n)$ curve shows the part where the possible equilibrium city size (with $n^e > \hat{n}$) is smaller than the optimum size. However, the equilibrium city size may also be larger than n^*.

As Prop. 2 makes clear, in the case of global emissions whether cities are over- or undersized depends on how per-capita emissions change with city population. However, as already stated in Section 3, not much is known about this relationship. Therefore, we estimate this relationship in Section 6, where we use numerical simulation to gauge whether cities will be over- or undersized in equilibrium.

5 Asymmetric cities

5.1 Equilibrium and social optimum with asymmetric city sizes

We now introduce asymmetric cities into the model. To do so, we assume that an individual living in city i obtains utility

$$v_i(n_i) = A_i n_i^{\gamma} (r_A + t n_i)^{-\alpha} E_i^{-\beta}. \tag{13}$$

[10]The functions have been rescaled so that $v(n^*) = \hat{v}(n^*)$ for better visibility.

The variable A_i is a city level amenity, which could be a consumption amenity such as good weather or a production amenity such as good infrastructure or a favourable geographic location. Without loss of generality, we assume $A_1 = 1$ and $A_i > A_{i+1}$ for $i = 1, 2, \ldots, m-1$.

As before, pollution is given by

$$E_i(\mathbf{n}) = e(n_i) + \delta \sum_{j \neq i} e(n_j),$$

with $e = n^\theta$, $\theta > 0$.

We assume the number of cities m is fixed and then ask how the optimum allocation of population among these cities differs from the equilibrium one.[11] Let $\hat{v}$ be the equilibrium utility level that is attained under free migration and let the equilibrium population vector be $\hat{\mathbf{n}} = \{\hat{n_1}, \ldots, \hat{n_m}\}$. The equilibrium city size distribution satisfies $v_i(\hat{\mathbf{n}}) = v_1(\hat{\mathbf{n}}) = v$ for all $i = 1, \ldots, m$. Using (13) and setting $A_1 = 1$, we can then solve for the amenity levels that are compatible with a free migration equilibrium:

$$A_i = \left(\frac{\hat{n}_1}{\hat{n}_i}\right)^\gamma \left(\frac{r_A + t\hat{n}_i}{r_A + t\hat{n}_1}\right)^\alpha \left(\frac{E_i(\hat{\mathbf{n}})}{E_1(\hat{\mathbf{n}})}\right)^\beta. \tag{14}$$

Note that our formulation implies that the amenity levels are uniquely identified by the equilibrium distribution of population up to the normalization that $A_1 = 1$.

We want to compare the equilibrium city size distribution to the optimal distribution. To characterize the latter, we assume the social planner maximizes the sum of utilities

$$\max_{\mathbf{n}} \sum_{i=1}^{m} n_i v_i(\mathbf{n})$$

subject to the population constraint $\sum_{i=1}^m n_i = N$. Letting λ be the Lagrangean multiplier on the population constraint, the first order conditions are given by[12]

$$v_i + n_i \frac{\partial v_i}{\partial n_i} + \sum_{j \neq i} n_j \frac{\partial v_j}{\partial n_i} = \lambda, \quad i = 1, \ldots, m. \tag{15}$$

[11] This differs from, e.g., Albouy et al. (2016) who study a city system with an endogenous number of asymmetric cities. However, looking at a varying number of cities in our context may not be reasonable. When we find that big cities are too small, we might want to close some small cities. Conversely, when big cities are too large, we might want to add more small cities. However, both exercises are not possible to implement in our numerical model, since we do not have data on the universe of all cities, so there is no 'smallest' city.

[12] We assume an interior solution where $0 < n_i^* < N$ for all i and that the second order conditions hold.

The last term on the LHS of (15) shows the pollution spillovers between cities.

The sign of $v_i - v_{i+1}$ is important in the following analysis. While $v_i = v_{i+1}$ holds at the equilibrium, suppose $v_i > v_{i+1}$ holds for all i at the social optimum. Because $\partial v_i / \partial n_i$ is negative in the neighborhood of a stable equilibrium for almost all i,[13] it must be that the optimum n_i is smaller than the equilibrium n_i in large cities, whereas the optimum n_i is larger than the equilibrium n_i in small cities. The opposite is true when $v_i < v_{i+1}$. Therefore, we can now show the following:

Proposition 3 *Assume that $\theta < 1$. Then (i) if pollution is close to local, the optimal utility is higher in larger cities. Large cities are too large and small cities are too small at the equilibrium; (ii) if pollution is close to global and the marginal damage of pollution is sufficiently large, the optimal utility is lower in larger cities. Large cities are too small and small cities are too large at the equilibrium.*

Proof. See Appendix A. ■

The intuition is as follows. Suppose that pollution is local, as might be the case, say, for NO_X. Then, the indirect utility v_i is a function of its city size n_i only. As shown by Henderson (1974), the indirect utility is decreasing in n_i at a stable equilibrium.

Start from the equilibrium $v_i = v_{i+1}$ with $n_i > n_{i+1}$ and consider the effect of moving one person from the larger city i to the smaller city $i + 1$. The utility v_i rises to $v_i + \Delta_i$ whereas the utility v_{i+1} falls to $v_{i+1} - \Delta_{i+1}$ because v_i decreases with n_i. The rise Δ_i and fall Δ_{i+1} are similar in magnitude when the one person is sufficiently small relative to total city size. Since there are more people in city i, however, the sum of $n_i\Delta_i$ exceeds the sum of $n_{i+1}\Delta_{i+1}$. Therefore, it is optimal to reduce the size of larger cities and raise that of smaller cities. As a result, the utility levels in larger cities are higher than those in smaller cities at the optimum.

By contrast, if pollution is global, such as in the case of CO_2, concentrating population in bigger cities decreases total emissions if $\theta < 1$, which benefits residents in all cities. When moving one person from a smaller city $i + 1$ to a larger city i, utility of city i residents falls while that of $i + 1$ residents rises. However, due to the global externality, utility of the residents of all other cities also rises. Therefore, as long as the marginal damage of pollution is large enough, social welfare rises.

Examining (15) in Appendix A, we can further say the following. Given sufficiently large δ (i.e., close to global pollution), large cities are more likely to be too small if the

[13] Tabuchi and Zeng (2004) show that a stable equilibrium requires $\partial v_i / \partial n_i < 0$ for at least $m - 1$ cities.

housing expenditure share α, the agricultural land rent r_A, and the commuting cost t are small. In this case, the crowding effects induced by commuting and tight housing markets in larger cities are outweighed by the beneficial effect of reduced pollution for all other cities.

In order to correct the discrepancy between the equilibrium and optimal distributions of city sizes, the national government may impose location taxes and subsidies according to city size. In the case of global pollution with $\theta < 1$, in our setup, living in large cities should be subsidized to make them more attractive.[14] We compute the optimal tax/subsidy scheme that achieves the socially optimal city size distribution under free mobility numerically below in Section 6.2.

6 Numerical simulation

6.1 Parameter values

We now try to assess to what extent optimum and equilibrium city size may diverge, using numerical simulation. We present here the results for the asymmetric city case. The symmetric city case can be found in Borck and Tabuchi (2016).

We use the following parameter values. We set the expenditure share of housing to $\alpha = 0.24$ following Davis and Ortalo-Magné (2011), and the agglomeration elasticity to $\gamma = 0.05$ (see Combes and Gobillon, 2015, for an overview).[15] From Borck and Brueckner (2016), we set $r_A = \$58,800$, the annual land rent of agricultural land in the US, and $t = \$503$, the annual (monetary plus time) commuting cost per mile in the US.

As described in Appendix C, we calibrate β, using central estimates of the social cost of carbon (SCC) from the literature. Using the central estimate of USD 40.54 per metric ton CO_2 for 2015 (assuming 3% discounting, value updated to 2015 USD) from the recent study by Interagency Working Group on Social Cost of Carbon (2015), we find a value of $\beta = 0.022$.[16,17]

[14] See also Eeckhout and Guner (2017) for an analysis of taxes related to city size.

[15] On the one hand, recent papers have found values of γ lower than 0.05 (Combes and Gobillon, 2015), on the other hand, accounting for dynamic externalities (De La Roca and Puga, 2017) or consumption benefits (Glaeser et al., 2001; Tabuchi and Yoshida, 2000), a value of 0.05 or higher may seem reasonable. We use both higher and lower values of γ later in our sensitivity checks.

[16] The study reports averages over three different integrated assessment models. We use their average value across the three models for 2015 at 3% discounting, USD 36 as our central value.

[17] In Appendix C, we calibrate β to match the lower SCC value for the US instead of the SCC for the world used by Interagency Working Group on Social Cost of Carbon (2015). Interestingly, since US

The estimates of the SCC are surrounded by a lot of uncertainty and some controversy. Therefore, we also use a higher estimate for the SCC. In particular, we use the 95th percentile estimate of USD 118 for 2015 from Interagency Working Group on Social Cost of Carbon (2015) (again at 3% discounting, in 2015 USD), which gives a value of $\beta = 0.064$. Finally, estimates for the social cost of carbon increase over time. For 2050, the central and 95th percentile values from Interagency Working Group on Social Cost of Carbon (2015) are USD 77.70 and 238.74. The implied values for β are 0.042 and 0.13.

Since much less is known about the emissions intensity θ than about the other parameters (see our discussion in Section 3), we estimate this parameter using US city data. Suppose that total emissions in city i in year t are $e_{it} = Bn_{it}^{\theta}$. Then, per capita emissions decrease with population size if and only if $\theta < 1$.

We can then estimate a linear regression of the form

$$\log e_{it} = c + \theta \log n_{it} + \varepsilon_{it} \tag{16}$$

where $c \equiv \log B$ is a constant, and ε is the error term.

We use data from Fragkias et al. (2013) to estimate CO_2-emissions in US core based statistical areas (metropolitan statistical areas and micropolitan areas) from 1999-2008. The dataset contains CO_2 emissions and population for 933 core based statistical areas (CBSAs). Emissions are based on data from the Vulcan Project, which quantifies U.S. fossil fuel carbon dioxide emissions at 10 km $\times$ 10 km grid cells and at the scale of individual factories, power plants, roadways and neighborhoods on an hourly basis. These are aggregated by Fragkias et al. (2013) to annual observations by CBSA.

Tab. 1 shows the summary statistics. Fig. 4 displays a binned scatter plot of log emissions against log population, where all data are pooled. Population varies from 12340 in the smallest city to 18.7 mill. in the New York metro area. Per capita emissions vary by a factor of over 200.

We start by estimating (16) by pooled OLS. Results are shown in column (1) of Tab. 2. Standard errors are clustered at the CBSA level. The coefficient on population is 0.938, and it is significantly smaller than one.[18] According to this estimate, if population doubles, per capita emissions would fall by $-(2^{\theta-1} - 1) \times 100 = 4.2\%$.[19]

population is smaller and per capita income higher, the resulting value of $\beta = 0.0266$ does not differ much from this baseline.

[18] Using Japanese data for 105 metropolitan employment areas in 2005, we obtain a similar result with a coefficient of 0.902, which is significantly smaller than one.

[19] To check for a possibly nonlinear relation between the log of emissions and log population, we also

Table 1: Summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Emissions (MMT)[a]	9330	1.542	3.924	.036	71.06
P.c. emissions (MMT)[a]	9330	8.4×10^{-6}	.000017	1.1×10^{-6}	.0002556
Population	9330	290029	994482.1	12340	1.87×10^{7}

[a] Million metric tons.
Source: Fragkias et al. (2013)

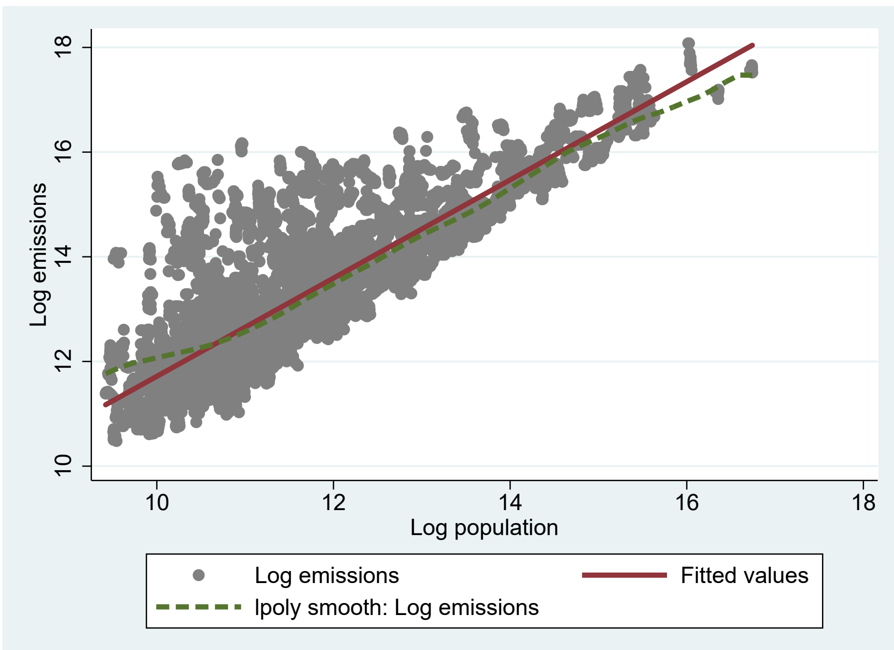

Figure 4: CO_2-emissions and city size

Note: The figure shows a scatterplot of CO_2 emissions by city size as well as a linear regression (solid line) and local polynomial smoothing (dashed)

Source: own calculations based on data from Fragkias et al. (2013)

Table 2: CO_2-emissions and city size

	dependent variable: log CO_2 emissions			
	(1)	(2)	(3)	(4)
Log population	0.938***	0.938***	0.834***	0.802***
	(0.0168)	(0.0168)	(0.0978)	(0.120)
Constant	2.335***	2.343***	3.533***	3.896***
	(0.202)	(0.201)	(1.117)	(1.369)
Observations	9,330	9,330	9,330	9,330
R-squared [within]	0.681	0.682	0.128	0.147
# of CBSAs	933	933	933	933
Year fixed effects	No	Yes	Yes	Yes
CBSA fixed effects	No	No	Yes	Yes
Division×Year fixed effects	No	No	No	Yes

Standard errors are clustered at the CBSA level.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Source: own calculations based on data from Fragkias et al. (2013).

This estimate may be biased due omitted variables or reverse causality. If pollution were local, then our model would predict that individual migration decisions are based on city emissions, so population would be endogenous and OLS estimation would consequently be biased. Given that CO_2 is a global pollutant, however, this is not a concern in the present setup, since migration should be independent of local emissions. Therefore, reverse causality may not be a big concern in the current setup. However, cities may still differ in unobserved factors that affect population size and emissions. To mitigate potential biases, we will add various fixed effects to our baseline regression.

First, in column (2), we include time fixed effects to allow for any time varying factors that are common across CBSAs and affect emissions, such as national business cycles. If these cycles were correlated with population size (say because some cities grow more than others when the economy grows) and also affect CO_2 emissions, the OLS coefficient would be biased. The coefficient in column (2), however, is the same as in the model without time effects.

include the square of log population. When we do so, the coefficient on log population drops slightly to 0.9, but the coefficient on the square term is insignificant. Fig. 4 compares a linear regression to local polynomial smoothing. The latter shows that the relation between log emissions and log population may be slightly nonlinear, but this nonlinearity seems restricted to the upper and lower tail of the city size distsribution.

In column (3) we include CBSA fixed effects. Some cities may have disproportionately many power plants that service larger geographic areas. Also, cities may differ in some unobserved dimension such as industry structure, climate, or other factors that may affect population size and emissions at the same time. As long as this heterogeneity is time invariant, we can control for it by estimating a model with CBSA fixed effects. As shown in column (3), the coefficient on population drops to 0.83 once we control for CBSA and time fixed effects. Thus, the result that $\theta < 1$ does not seem to be driven by unobserved heterogeneity among CBSAs.

Finally, there may be macroeconomic effects that affect regions differentially and may be correlated with city size and emissions. For instance, cities in industrial regions and those in regions with a large service sector will be differentially affected by business cycles, and so will their CO_2 emissions. To deal with time varying regional differences of this sort, in column (4), we include interaction effects between year and US census divisions (there are 4 census regions and 9 divisions). As can be seen from the Table, the coefficient on population slightly drops to 0.8, and it remains significantly smaller than one.[20] Hence, doubling population would reduce per capita CO_2 emissions by $-(2^{0.8-1} - 1) \times 100 = 12.8\%$. This is our preferred estimate, since it controls extensively for time-varying regional heterogeneity, and we will use this value of θ for the numerical simulation. However, we will also use the higher value of 0.94 as a robustness check.[21]

6.2 Simulation results

We now simulate numerically the equilibrium and optimal city size distribution with asymmetric cities. We assume a given number of cities, m, and given total population N. So we exclude the formation of new cities. For our simulation, we use the 180 largest US CBSAs. The total population is the sum of the population sizes of these 180 cities,

[20]The p-values for the test for $\theta < 1$ are: 0.0001 in column (1), 0.0001 in column (2), 0.045 in column (3) and .0499 in column (4). In all cases, the hypothesis that $\theta < 1$ cannot be rejected at the 5% significance level.

[21]An interesting question is whether the effect we estimate is due to population size or population density, since the two may affect emissions differently. In Appendix Table A.1, we add the log of population density as a regressor. In the OLS regressions (col. 1 and 2 of Tab. A.1), the coefficient on population increases to 1.04, while the coefficient on density is -0.16. On this count, one could argue that population density, rather than population size, drives the result that population increases emissions less than proportionately. In the fixed effects regressions (cols. 3 and 4 of Tab. A.1), however, the coefficient on population drops to 0.6-0.65 once we add density, while the coefficient on density is positive, although not significant. On this count, one could argue that density does not influence emissions, while population size increases emissions less than proportionately. However, since our model predicts that density is itself a function of population, the results should be interpreted with caution.

$n = 225,678,243$. In Appendix D, we present an alternative simulation where we assume 180 cities with the same total population and a city size distribution which follows Zipf's law. We then compare the equilibrium number of cities to the social optimum.

Results. Out of all CBSAs in the year 2008, we keep the largest 180 cities (see Tab. 3 for the ten largest and smallest CBSAs).[22] These cities comprise 90% of the total population living in MSAs (252 mill.) and 80% of the population in CBSAs (283 mill.).

Table 3: The ten largest and smallest MSAs in the simulation

Rank	MSA	Amenity	Population	Δ population	Emissions	Δ emissions
1	New York-Northern New Jersey-Long Island, NY-NJ-PA	1	18672355	4.18 %	40.7	3.33 %
2	Los Angeles-Long Beach-Santa Ana, CA	0.929	12692740	3.55 %	24.5	2.83 %
3	Chicago-Joliet-Naperville, IL-IN-WI	0.877	9384555	3.02 %	42.6	2.41 %
4	Dallas-Fort Worth-Arlington, TX	0.81	6158022	2.24 %	17.5	1.79 %
5	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	0.804	5906917	2.16 %	15.9	1.72 %
6	Houston-Sugar Land-Baytown, TX	0.798	5702270	2.09 %	24.	1.67 %
7	Miami-Fort Lauderdale-Pompano Beach, FL	0.792	5454633	2. %	12.7	1.6 %
8	Washington-Arlington-Alexandria, DC-VA-MD-WV	0.79	5391607	1.98 %	19.3	1.58 %
9	Atlanta-Sandy Springs-Marietta, GA	0.783	5152141	1.89 %	25.4	1.51 %
10	Boston-Cambridge-Quincy, MA-NH	0.763	4483141	1.61 %	13.7	1.28 %
171	Clarksville, TN-KY	0.444	261530	-5.94 %	4.9	-4.78 %
172	Myrtle Beach-North Myrtle Beach-Conway, SC	0.444	260609	-5.95 %	1.01	-4.79 %
173	Santa Cruz-Watsonville, CA	0.443	256520	-6. %	0.626	-4.83 %
174	Cedar Rapids, IA	0.443	255503	-6.02 %	0.92	-4.84 %
175	Binghamton, NY	0.442	252527	-6.06 %	0.869	-4.88 %
176	Merced, CA	0.441	250538	-6.09 %	0.747	-4.9 %
177	Lynchburg, VA	0.44	249299	-6.1 %	0.972	-4.91 %
178	Bremerton-Silverdale, WA	0.44	246912	-6.13 %	0.4	-4.94 %
179	Amarillo, TX	0.439	244454	-6.17 %	3.3	-4.96 %
180	Olympia, WA	0.439	244332	-6.17 %	0.467	-4.97 %

Note: The table displays population and emissions levels by MSA for 2008 from Fragkias et al. (2013).
Amenity levels are computed from (14); Δ refers to simulated population and emissions changes at optimum as described in the text.

We assume that the current distribution is an equilibrium. We can then compute the level of amenities that rationalize the equilibrium from (14) (the values are in Tab. 3 for the ten largest and ten smallest cities, and in Tab. OA.1 in the Online Appendix for all cities). Fig. 5 shows the equilibrium (light blue/gray) and optimal (dark red/gray) city size distribution. For better visibility, the figure plots the equilibrium and optimal distributions assuming the larger β value of 0.064. Zipf's law holds fairly well for the upper tail of the distribution. Note that the largest city in the sample, New York, has 18.7 mill. whereas

[22] The full sample of cities is displayed in Tab. OA.1 in the online Appendix.

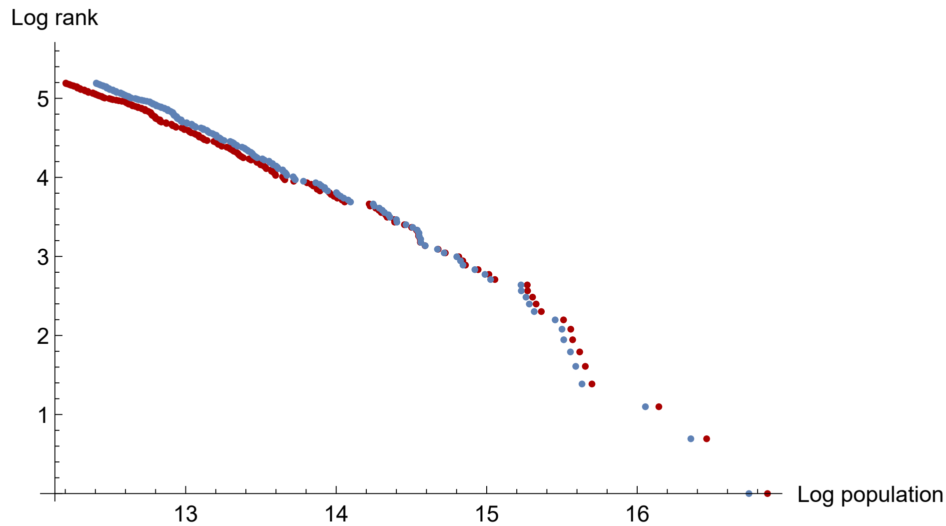

Figure 5: Optimal (dark red/grey) and equilibrium city size distributions (light blue/gray)
Source: own calculations based on data from Fragkias et al. (2013)

according to Zipf's law the largest city has more than twice that many inhabitants. Tab. 3 shows the equilibrium population level, emission level, as well as the percentage change rates of the optimum relative to equilibrium levels (denoted by Δ).

As shown in Fig. 5 and Tab. 3, the three largest cities are undersized by 3-4%. At the social optimum, the largest city in the sample, New York, is undersized by 779,843 of its 18.7 mill. inhabitants, while the second largest, Los Angeles, is undersized by 450,180. The smallest city in the sample is oversized by about 6.2%. Note that out of the 180 cities, 23 are undersized and the other 157 are oversized at the equilibrium. In total, moving from the equilibrium to the optimal allocation would require moving 5.35 mill. people or 2.4% of the total population.

The table also shows the changes in emissions, which amount to 80% of the changes in population levels. Total emissions fall from 3.74×10^8 MMT to 3.72×10^8, or by 0.6%.

Mirroring the population changes, Tab. 4 shows changes in wages and housing rents.[23] At the social optimum, the largest cities command higher wages and rents, but as shown in Prop. 3, at the optimum, the large city residents receive lower utility than small city residents.

[23]The wage rate in city i is assumed to be Bn_i^γ, where we set B to target the average wage in the US in 2008, \$52,029. In order to compute rents, we choose units of housing to match average dwelling size in the US, 2,196 sq ft. (see Borck and Brueckner (2016) for details).

Table 4: The ten largest and smallest MSAs in the simulation (2)

Rank	MSA	Population	Wage	Δ wage	Housing rent	Δ housing rent
1	New York-Northern New Jersey-Long Island, NY-NJ-PA	18672355	61165.1	0.20 %	23.3157	4.18 %
2	Los Angeles-Long Beach-Santa Ana, CA	12692740	59995.9	0.17 %	15.8491	3.55 %
3	Chicago-Joliet-Naperville, IL-IN-WI	9384555	59096.8	0.15 %	11.7183	3.02 %
4	Dallas-Fort Worth-Arlington, TX	6158022	57865.	0.11 %	7.68946	2.24 %
5	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	5906917	57744.6	0.11 %	7.37592	2.16 %
6	Houston-Sugar Land-Baytown, TX	5702270	57642.9	0.10 %	7.12038	2.09 %
7	Miami-Fort Lauderdale-Pompano Beach, FL	5454633	57515.1	0.10 %	6.81117	2.00 %
8	Washington-Arlington-Alexandria, DC-VA-MD-WV	5391607	57481.7	0.10 %	6.73247	1.98 %
9	Atlanta-Sandy Springs-Marietta, GA	5152141	57351.3	0.09 %	6.43345	1.89 %
10	Boston-Cambridge-Quincy, MA-NH	4483141	56953.8	0.08 %	5.5981	1.61 %
171	Clarksville, TN-KY	261530	49410.5	-0.31 %	0.32671	-5.94 %
172	Myrtle Beach-North Myrtle Beach-Conway, SC	260609	49401.8	-0.31 %	0.32556	-5.94 %
173	Santa Cruz-Watsonville, CA	256520	49362.8	-0.31 %	0.320454	-6.00 %
174	Cedar Rapids, IA	255503	49353.	-0.31 %	0.319184	-6.02 %
175	Binghamton, NY	252527	49324.1	-0.31 %	0.315468	-6.06 %
176	Merced, CA	250538	49304.6	-0.31 %	0.312984	-6.08 %
177	Lynchburg, VA	249299	49292.4	-0.31 %	0.311437	-6.10 %
178	Bremerton-Silverdale, WA	246912	49268.7	-0.32 %	0.308457	-6.13 %
179	Amarillo, TX	244454	49244.	-0.32 %	0.305388	-6.17 %
180	Olympia, WA	244332	49242.8	-0.32%	0.305235	-6.17 %

Note: Δ refers to simulated wage and rent changes at optimum as described in the text.

The welfare gain from moving to the optimal city size distribution is small, less than 0.1% of income.[24] Small welfare gains from optimal policies are also found in similar models e.g. by Eeckhout and Guner (2017) and Albouy et al. (2016).

We can also compute the level of taxes that could be used by a social planner to support the social optimum. To do so, let the utility level in city i be $v(n_i, \tau_i)$, where τ_i is a city specific tax or subsidy. Denoting optimal population levels by n_i^*, the Pigouvian taxes/subsidies τ_i solve the linear system of equations

$$v(n_i^*, \tau_i) = v(n_j^*, \tau_j) \tag{17}$$

$$\sum_{j=1}^{m} \tau_j = 0, \quad \text{for } i, j = 1, ..., m, \ i \neq j \tag{18}$$

When the optimal tax/subsidy scheme is applied, free migration will lead to the city size distribution that replicates the social optimum.

[24]When population is efficiently allocated, as noted, total emissions fall by 0.6%, relative to the equilibrium. Note, however, that the welfare gain from efficiently allocating population in the absence of pollution would also be small with our parameters.

The computation reveals that the optimum is supported by a subsidy on the 72 largest cities and taxes on the rest of the cities. Living in the largest city, New York, would have to be supported by a subsidy corresponding to 1.4% of the local wage. The tax levels for the small cities are all below one percent of the local wage.

Finally, note that we have assumed that individuals are freely mobile. One could study a model with imperfect mobility, for instance by assuming heterogeneous location preferences. Depending on the distribution of these preferences, a social planner would want to move fewer individuals than in the absence of moving costs.

Sensitivity. We now briefly describe how the results change when we vary some of the model parameters. First, we increase θ to 0.94 to reflect a potentially higher pollution elasticity. We find that the largest city is undersized by 1.6% and the smallest oversized by 2%. Conversely, when θ decreases to 0.75, the biggest city is undersized by 4.7% and the smallest city is oversized by 7.5%. Since the benefit of concentration is increased the more per capita pollution decreases with city size, this finding is intuitive.

Next, suppose that in line with a high estimate of the social cost of carbon, the emissions damage, β, increases from 0.022 to 0.064 (see Section 6). Now, the largest cities are undersized by 9-13% and the smallest oversized by 18%.[25] When the agglomeration elasticity, γ, increases from 0.05 to 0.08, agglomeration becomes more efficient, and again, concentration increases at the optimum: the largest city is undersized by 4.8% at the equilibrium and the smallest oversized by 7%.[26] Conversely, when we decrease γ to a lower value of 0.02, the largest city is undersized by 3.7% and the smallest one oversized by 5.5%.

Lastly, we change the spillover parameter δ. As a first variation, we use completely local emissions, that is, $\delta = 0$. As shown in Prop. 3, the biggest cities will then be oversized and the smallest cities undersized. However, our computations reveal small overall differences between optimum and equilibrium city sizes. Therefore, we conclude that when pollution is completely local, big cities are only slightly too big and the equilibrium city size distribution is close to optimal.

A natural question to ask is, what is the actual degree of spillovers from local emissions? To approach this question, we borrow from Borck and Brueckner (2016) who consider optimal energy taxation with local and global emissions. Their results imply that local

[25] For completeness, we also report results for a low SCC value. When β is 0.012 (corresponding to a low SCC value of USD 20), the largest cities are undersized by around 2 percent and the smallest oversized by 3.3 percent.

[26] In fact, Tabuchi and Yoshida (2000) show that agglomeration externalities from consumption in Japanese cities are about the same size as productive externalities.

emissions make up 60.7% of emissions from commuting and 53% of emissions from residential energy use.[27] Let us take the average of these values, 57%, so we set $\delta = 1 - 0.57 = 0.43$. We then find that the largest city is undersized by 3.5% and the smallest is oversized by 5.8%. In the past two decades local pollution has decreased relative to global pollution and this trend is likely to continue (see, e.g., Amann et al., 2013). Moreover, the damage from global warming is projected to increase over time, since GHGs accumulate in the atmosphere and warming is caused by the stock of pollution. Therefore, we tentatively conclude that using realistic parameters, the case for large cities being undersized and small ones oversized persists and will get stronger over time.

On the other hand, increasing agricultural land rent r_A to $100,000 per year increases the costs of agglomeration. The effect on the optimum size of cities is rather small, however. A very similar result obtains when the per mile commuting cost increases to $750 per year.

7 Conclusion

The paper has analyzed the optimum size of cities in an urban model with environmental pollution. When pollution is purely local and cities are symmetric, we find that equilibrium cities are too large, mirroring the finding of Henderson (1974) and others. With asymmetric cities, this translates into the result that big cities are oversized and small cities undersized.

However, when pollution is global and per capita pollution decreases in population size, we find that in a symmetric city model, cities might be inefficiently small, contrary to the standard model. When cities are asymmetric, big cities are undersized and small cities oversized. Over the last decades, global pollution has increased relative to local pollution, and the damage from global warming increases over time. Hence, we conclude that for the future, a policy which favors big cities might actually be warranted.

Some possibilities for future research suggest themselves. First, our analysis was based on one estimate of the population elasticity of pollution, which is a central parameter in the analysis. More robust evidence on this parameter clearly seems important. Second, we think it would be interesting to redo the quantitative analysis with data from different countries. For instance, there is a growing number of papers on Chinese cities (e.g. Au and Henderson, 2006). Since the properties of the equilibrium city system and pollution patterns in China and other developing economies is undoubtedly different from developed

[27] It is assumed that local pollution is measured in units such that the same marginal damage value can be applied to global and local pollution.

countries, studying equilibrium and optimum city systems in this context would seem to be relevant.

Finally, we could include city governments that maximize residents' utility. We conjecture that a city government would want a city size even smaller than the free mobility case. The reason is that in the symmetric case, the free mobility outcome is at least the n which maximizes $\hat{v}(n)$, the utility without pollution, whereas a city government would maximize roughly the utility $\tilde{v}(n)$ which would correspond to $\hat{v}(n)$ plus the disutility from local pollution, since the local government cares only for the locally produced pollution affecting its own citizens.[28] This could magnify the welfare gain from imposing optimal city sizes.

[28]This assumes global pollution. With local pollution, a city government maximizes $v(n)$ so the allocation would be efficient in the symmetric case. See, e.g. Abdel-Rahman and Anas (2004) and the literature therein.

Appendix

A Proof of Proposition 3

Since

$$
\begin{aligned}
n_i \frac{\partial v_i}{\partial n_i} &= \gamma v_i - \frac{\alpha t n_i}{r_A + t n_i} v_i - \beta \theta n_i^{\theta} v_i E_i^{-1} \\
\sum_{j \neq i} n_j \frac{\partial v_j}{\partial n_i} &= -\beta \theta \delta n_i^{\theta - 1} \sum_{j \neq i} n_j v_j E_j^{-1},
\end{aligned}
$$

we have

$$
\begin{aligned}
& v_i + n_i \frac{\partial v_i}{\partial n_i} + \sum_{j \neq i} n_j \frac{\partial v_j}{\partial n_i} - \lambda \\
& = \left(1 + \gamma - \frac{\alpha t n_i}{r_A + t n_i}\right) v_i - \beta \left(1 - \delta\right) \theta v_i n_i^{\theta} E_i^{-1} - \beta \delta \theta Z n_i^{\theta - 1} - \lambda \\
& = X_i v_i - \beta \delta \theta Z n_i^{\theta - 1} - \lambda \qquad \text{(A.1)} \\
& = 0,
\end{aligned}
$$

where

$$
X_i \equiv 1 + \gamma - \frac{\alpha t n_i}{r_A + t n_i} - \beta \left(1 - \delta\right) \theta n_i^{\theta} E_i^{-1}
$$

and $Z \equiv \sum_j n_j v_j E_j^{-1}$ is constant across cities.

Since the expression (A.1) is the same for i and for $i+1$, we can eliminate λ as follows:

$$
X_i v_i - \beta \delta \theta Z n_i^{\theta - 1} = X_{i+1} v_{i+1} - \beta \delta \theta Z n_{i+1}^{\theta - 1},
$$

which can be rewritten as

$$
v_i = \frac{1}{X_i} \left(X_{i+1} v_{i+1} + \beta \delta \theta Z n_i^{\theta - 1} - \beta \delta \theta Z n_{i+1}^{\theta - 1}\right).
$$

Thus, the utility differential is

$$\begin{aligned}\Delta v &\equiv v_i - v_{i+1} \\ &= \frac{1}{X_i}\left[\left(X_{i+1} - X_i\right) v_{i+1} + \beta\delta\theta Z\left(n_i^{\theta-1} - n_{i+1}^{\theta-1}\right)\right] \\ &= \Delta V_a + \Delta V_b,\end{aligned}$$

where

$$\begin{aligned}\Delta V_a &\equiv \frac{\alpha t r_A v_{i+1}\left(n_i - n_{i+1}\right)}{\left(r_A + t n_i\right)\left(r_A + t n_{i+1}\right)} \\ \Delta V_b &\equiv \beta\theta\left[\left(1-\delta\right) v_{i+1}\left(n_i^{\theta} E_i^{-1} - n_{i+1}^{\theta} E_{i+1}^{-1}\right) + \delta Z\left(n_i^{\theta-1} - n_{i+1}^{\theta-1}\right)\right].\end{aligned}$$

While $\Delta V_a > 0$, the sign of ΔV_b is indeterminate. However, the first term of ΔV_b is positive whereas the second term of ΔV_b is negative because

$$\begin{aligned}n_i^{\theta} E_i^{-1} - n_{i+1}^{\theta} E_{i+1}^{-1} &= \frac{1}{E_i E_{i+1}}\left(n_i^{\theta} E_{i+1} - n_{i+1}^{\theta} E_i\right) \\ &= \frac{\delta}{E_i E_{i+1}}\left(n_i^{\theta} - n_{i+1}^{\theta}\right)\sum_j n_j^{\theta} > 0\end{aligned}$$

and

$$n_i^{\theta-1} - n_{i+1}^{\theta-1} < 0, \ \forall\theta \in (0,1).$$

(i) Let $\delta = 0$. Then, $\Delta V_b > 0$, and thus $\Delta v > 0$. By continuity, this also holds for δ positive but close to zero.

(ii) Let $\delta = 1$. Solving $\Delta v < 0$ for β, we have

$$\Delta v < 0 \Leftrightarrow \beta > \tilde{\beta} \equiv \frac{\alpha t r_A v_{i+1}(n_{i+1} - n_i)}{\theta Z(r_A + t n_i)(r_A + t n_{i+1})(n_i^{\theta-1} - n_{i+1}^{\theta-1})} > 0.$$

By continuity, $\Delta v < 0$ holds for sufficiently large β when δ is close to but smaller than one. $\blacksquare$

B Local landownership

Suppose that all land in a city is owned by residents, so the total differential land rent is distributed equally to all residents. Let income be given by $y = w + R/n$, where

$$R = \int_0^{\bar{x}} (r(x, v) - r_A)dx \tag{A.2}$$

is the total differential land rent. Rewriting (8) and (7) gives

$$\bar{x} = \frac{(w + R/n)\left[1 - r_A^\alpha (r_A + tn)^{-\alpha}\right]}{t} \tag{A.3}$$

$$v = (w + R/n)(r_A + tn)^{-\alpha} E^{-\beta}. \tag{A.4}$$

Substituting from (A.4) into $r(x, v) = (w + R/n - tx)^{1/\alpha} E^{-\beta/\alpha} v^{-1/\alpha}$ with $w = n^\gamma$ gives $r(x, v) = (r_A + tn)\left(n^\gamma + R/n\right)^{-1/\alpha}\left(n^\gamma + R/n - tx\right)^{\frac{1}{a}}$. Using this in (A.2) and solving gives

$$R = \frac{n^{1+\gamma}\left[r_A^{1+\alpha} - (r_A + tn)^\alpha (r_A - \alpha tn)\right]}{(r_A + tn)^{1+\alpha} - r_A^{1+\alpha}}. \tag{A.5}$$

Finally, substituting in (A.4) gives

$$v = \frac{(1 + \alpha)tE^{-\beta} n^{1+\gamma}}{(r_A + tn)^{1+\alpha} - r_A^{1+\alpha}}. \tag{A.6}$$

which is also inverted U-shaped in n.

We then redo the simulation exercise from Section 6.2. For the city size distribution described by Zipf's law with $\delta = 1$, we find the largest city is undersized by 3.6% and the smallest is oversized by 7.3%, so results are very close to the baseline simulation. Varying δ shows that this also holds for local pollution and for the intermediate case $\delta = 0.43$.

C Calibration of β

We now calibrate β using central estimates of the social cost of carbon from Interagency Working Group on Social Cost of Carbon (2015). The MRS between pollution and (non-housing) consumption is

$$\text{MRS} = -\frac{\partial u/\partial E}{\partial u/\partial z} = \frac{\beta z}{(1 - \alpha)E}. \tag{A.7}$$

Substituting optimal consumption, $z(y) = (1-\alpha)(w-tx)$ gives MRS $= \beta(w-tx)/E$, and integrating over the city gives citywide MRS

$$\overline{\text{MRS}} = \int_0^{\bar{x}} \frac{\beta(w-tx)}{E} \frac{1}{s(x)} dx \tag{A.8}$$

$$= \frac{\beta w \left[tn + r_A - r_A^{1+\alpha}(tn+r_A)^{-\alpha}\right]}{(1+\alpha)tE}. \tag{A.9}$$

where we have substituted the optimal $s(x)$ and used (7) and (8). Finally, letting M be world population and n be city population, we get the social cost of carbon

$$\text{SCC} = \frac{\beta M w \left[tn + r_A - r_A^{1+\alpha}(tn+r_A)^{-\alpha}\right]}{(1+\alpha)tnE}. \tag{A.10}$$

We use the following parameters: world population in 2015 was $M = 7.35$ billion (source: UN World Population Prospects, `http://esa.un.org/unpd/wpp/Download/Standard/Population/`), world per capita income in 2015 was $w = \$10,743$ (source: UN National Accounts Main Aggregates Database, `http://unstats.un.org/unsd/snaama/dnllist.asp`), and total CO_2 emissions were $E = 34,649$ million metric tons CO_2 in 2011 (source: World Bank, World Development Indicators, `http://data.worldbank.org/indicator/EN.ATM.CO2E.KT/countries`). We set $n = 750,000$ and from Borck and Brueckner (2016), we use $t = \$503.53, \alpha = 0.24$, and $r_A = \$58,800$.

The target value for the social cost of carbon is USD 40.54 per ton CO_2, using the central value from Interagency Working Group on Social Cost of Carbon (2015) (converted from 2007 to 2015 USD). Using the stated parameters, setting (A.10) equal to 40.54 and solving gives $\beta = 0.022$. The other values in the text are solved likewise using different values for the social cost of carbon.

Calibration using SCC for US. An alternative approach would be to calibrate β to target the SCC for the USA, on the assumption that national policy makers care about national welfare, not the welfare of world citizens. Therefore, we recalibrate β to target the SCC value for the US computed by Nordhaus (2017). His central US estimate is \$4.78, much lower than the \$31.21 value for the world. We therefore now use a target value of \$5 instead of the world value of \$40. Setting income per capita for the USA at \$56,516 and population at 324 mill. and solving (A.10) gives a value of $\beta = 0.0266$. Interestingly, therefore, the results do not differ much when we use an SCC value for the US instead of the world SCC.

D Simulation assuming Zipf's law.

In this exercise, we assume that the city size distribution follows Zipf's law. As is well known, this is a good approximation for city systems in most countries, except at the very top and bottom of the distribution (Gabaix, 2016). We continue to assume $n = 225,678,243$, and distribute this population to 180 cities so the city size distribution exactly follows Zipf's law. The largest city has 39 mill. inhabitants (more than twice the New York City metro area), the second largest 19.5 mill. and the smallest city has 217,180 inhabitants. We compute the amenity levels from (14) for these given population sizes.

We first assume $\delta = 1$ so pollution is global. We find that the biggest city is undersized by 4.8% (1.9 mill.) while the smallest one is oversized by 7.8%.

When pollution is purely local, $\delta = 0$, we find that the divergence between optimal and equilibrium city sizes is small, as in the real cities sample.[29] The largest city is oversized by 0.01% and the smallest city is undersized by 0.06%.

Finally, for an intermediate value, $\delta = 0.43$, the largest city is undersized 3.6% and the smallest oversized by 7%.

References

Abdel-Rahman, H.M. (1988). Product differentiation, monopolistic competition and city size. *Regional Science and Urban Economics* 18, 69–86.

Abdel-Rahman, H.M. and A. Anas (2004). Theories of systems of cities. In: J.V. Henderson and J.-F. Thisse (eds.), *Handbook of Regional and Urban Economics*, vol. 4, 2293–2339, Amsterdam: Elsevier.

Albouy, D., K. Behrens, F. Robert-Nicoud and N. Seegert (2016). The optimal distribution of population across cities. NBER Working Paper 22823.

Amann, M., Z. Klimont and F. Wagner (2013). Regional and global emissions of air pollu-

[29] We note that this is partly due to the assumption that the value of β is the same for local and global pollution, but local pollution is much smaller than global pollution. Therefore, for a more realistic simulation, the marginal damage of local emissions relative to global emissions should probably be increased.

tants: Recent trends and future scenarios. *Annual Review of Environment and Resources* 38, 31–55.

Arnott, R. Optimal city size in a spatial economy. *Journal of Urban Economics* 6, 65–89.

Au, C.C. and J.V. Henderson (2006). Are Chinese cities too small? *Review of Economic Studies* 73, 549–576.

Blaudin de Thé, C. and M. Lafourcade (2016). The carbon footprint of suburbanization: Evidence from French household data. Mimeographed.

Borck, R. (2016). Will skyscrapers save the planet? Building height limits and urban greenhouse gas emissions. *Regional Science and Urban Economics* 58, 13–25.

Borck, R. and J.K. Brueckner (2018). Optimal energy taxation in cities. *Journal of the Association of Environmental and Resource Economists 5, 481–516.*

Borck, R. and M. Pflüger (2015). Green cities? Urbanization, trade and the environment. IZA discussion paper 9104.

Borck, R. and T. Tabuchi (2016). Pollution and City Size: Can Cities Be Too Small? CESifo Working Paper No. 6152.

Combes, P.-P. and L. Gobillon (2015). The empirics of agglomeration economies. In G. Duranton, J. V. Henderson and W. C. Strange (eds.), *Handbook of Regional and Urban Economics* vol. 5, 247–348, Amsterdam: Elsevier.

Dascher, K. (2014). City Silhouette, World Climate. Available at `http://ssrn.com/abstract=2250673`.

Davis, M.A. and F. Ortalo-Magné (2011). Household expenditures, wages, rents. *Review of Economic Dynamics* 14, 248–261.

De La Roca, J. and D. Puga (2017). Learning by working in big cities. *Review of Economic Studies*, 84, 106–142.

Diamond, R. (2016). The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000. *American Economic Review* 106, 479–524.

Duranton, G. and D. Puga (2004). Micro-foundations of urban agglomeration economies, in J.V. Henderson and J.-F. Thisse (eds.), *Handbook of Regional and Urban Economics*, vol.4, 2063–2117, Amsterdam: Elsevier.

Eeckhout, J. and N. Guner (2017). Optimal spatial taxation: Are big cities too small? CEMFI working paper no. 1705.

Fragkias, M., J. Lobo, D. Strumsky and K.C. Seto (2013). Does size matter? Scaling of CO2 emissions and U.S. urban areas. *PLoS ONE* 8(6): e64727. doi: 10.1371/journal.pone.0064727

Gabaix, X. (2016). Power laws in economics: An introduction. *Journal of Economic Perspectives* 30, 185–206.

Gaigné, C., S. Riou and J.-F. Thisse (2012). Are compact cities environmentally friendly? *Journal of Urban Economics* 72, 123–136.

Glaeser, E. (2011). *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*, New York: Penguin Press.

Glaeser, E.L. and M.E. Kahn (2010). The greenness of cities: Carbon dioxide emissions and urban development. *Journal of Urban Economics* 67, 404–418.

Glaeser, E.L., Kolko, J. and A. Saiz (2001). Consumer city. *Journal of Economic Geography* 1, 27–50.

Gudipudi, R., T. Fluschnik, A. Garcìa Cantú Ros, C. Walther and J.P. Kropp (2016). City density and CO2 efficiency, *Energy Policy* 91, 352–361.

Henderson, J.V. (1974). The sizes and types of cities. *American Economic Review* 64, 640–656.

Hsieh, C.-T. and E. Moretti. (2017). Housing Constraints and Spatial Misallocation. NBER Working Paper 21154.

Interagency Working Group on Social Cost of Carbon, United States Government (2015). Technical Support Document: Technical Update of the Social Cost of Carbon for Regulatory Impact Analysis Under Executive Order 12866. US Government Printing Office, Washington D.C.

Lamsal, L.N., R.V. Martin, D.D. Parrish and N.A. Krotkov (2013). Scaling relationship for NO2 pollution and urban population size: A satellite perspective. *Environmental Science & Technology* 47, 7855–7861.

Larson, W. and A.M. Yezer (2015). The energy implications of city size and density. *Journal of Urban Economics* 90, 35–49.

Larson, W., F. Liu and A. Yezer (2012). Energy footprint of the city: Effects of urban land use and transportation policies. *Journal of Urban Economics* 72, 147–159.

Morikawa, M. (2013). Population density and efficiency in energy consumption: An empirical analysis of service establishments. *Energy Economics* 34, 1617–1622.

Nordhaus, W.D. (2017). Revisiting the social cost of carbon. *Proceedings of the National Academy of Sciences* 114, 1518–1523.

Redding, S.J. and Rossi-Hansberg, E.A. (2017). Quantitative spatial economics. *Annual Review of Economics* 9, 21–58.

Rybski, D., D.E. Reusser, A.-L. Winz, C. Fichtner, T. Sterzel and J.P. Kropp (2016). Cities as nuclei of sustainability? *Environment and Planning B*, first published on May 5, 2016, doi:10.1177/0265813516638340.

Sarzynski, A. (2012). Bigger is not always better: A comparative analysis of cities and their air pollution impact. *Urban Studies* 49, 3121–3138.

Seto, K.C., B. Güneralp and L.R. Hutyra (2012). Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences* 109, 16083–16088.

Tabuchi, T. and A. Yoshida (2000). Separating urban agglomeration economies in consumption and production. *Journal of Urban Economics* 48, 70–84.

Tabuchi, T. and D.-Z. Zeng (2004). Stability of spatial equilibrium. *Journal of Regional Science* 44, 641–660.

Tolley, G.S. (1974). The welfare economics of city bigness. *Journal of Urban Economics* 1, 324–345.

Tscharaktschiew, S. and G. Hirte (2010). The drawbacks and opportunities of carbon charges in metropolitan areas – a spatial general equilibrium approach. *Ecological Economics* 70, 339–357.

Appendix

Table A.1: CO_2-emissions and city size (2)

	Dependent variable: log CO_2 emissions			
	(1)	(2)	(3)	(4)
Log population	1.040***	1.040***	0.648***	0.599***
	(0.0281)	(0.0281)	(0.176)	(0.186)
Log density	-0.158***	-0.158***	0.181	0.200
	(0.0369)	(0.0369)	(0.147)	(0.142)
Constant	2.335***	2.343***	3.533***	3.896***
	(0.202)	(0.201)	(1.117)	(1.369)
Observations	9,244	9,244	9,244	9,244
R-squared [within]	0.689	0.691	0.128	0.147
# of CBSAs	928	928	928	928
Year fixed effects	No	Yes	Yes	Yes
CBSA fixed effects	No	No	Yes	Yes
Division×Year fixed effects	No	No	No	Yes

Standard errors are clustered at the CBSA level.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Source: own calculations based on data from Fragkias et al. (2013).